UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/804,724	03/19/2004	Jerry Rolia	200300271-1	8258

22879 7590 05/25/2010

HEWLETT-PACKARD COMPANY
Intellectual Property Administration
3404 E. Harmony Road
Mail Stop 35
FORT COLLINS, CO 80528

EXAMINER	
WILLIAMS, CLAYTON R	
ART UNIT	PAPER NUMBER
2457	

NOTIFICATION DATE	DELIVERY MODE
05/25/2010	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

JERRY.SHORMA@HP.COM
ipa.mail@hp.com
laura.m.clark@hp.com

Office Action Summary	Application No.	Applicant(s)	
	10/804,724	ROLIA ET AL.	
	Examiner	Art Unit	
	Clayton R. Williams	2457	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *12 February 2010*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1,2,4-12,14-16,18-26 and 28-30* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1,2,4-12,14-16,18-26 and 28-30* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1. Claims 1, 2, 4-12, 14-16, 18-26 and 28-30 are pending in this application per amendment.

Claim Rejections - 35 USC § 103

2. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

3. Claims 1, 2, 4-12, 14-16, 18-26 and 28-30 are rejected under 35 U.S.C. 102(e) as being anticipated by Gasca, Jr. et al (20050198231: hereinafter Gasca), in view of Shabauddin (6877035: hereinafter Shabauddin).

For claims 1, 15, 29 and 30, Gasca discloses:

A method of policing resources in a computing utility facility, comprising:

intercepting an advanced request for resources from an application admitted to access a pool of resources associated with the computing utility facility and prior to utilization of the pool of resources to execute the application ([0057], lines 1-4: “At block 300, the process waits for a new provisioning request”. The utility computing system receives a provisioning request on behalf a customer application.);

acquiring an entitlement profile associated with the application to determine if application is entitled to requested resources over a time period ([0057], lines 5-8: "At block 302, customer entitlement is checked. Checking customer entitlement (302) is a matter of knowing what kind of resources are promised in the service level agreement (SLA)". Furthermore, [0057], lines 16-18: "Additionally, there could be time constraints (e.g. the customer is entitled to 20 servers between 8 am and 5 pm, but only 10 servers the rest of the time.");

identifying an entitlement value and corresponding sliding window of the time period from the entitlement profile ([0057], lines 16-18: Utility computing system curtails use of resources with respect to specific time windows. "Additionally, there could be time constraints (e.g. the customer is entitled to 20 servers between 8 am and 5 pm, but only 10 servers the rest of the time.");

determining if the request for resources exceeds the entitlement value associated with the sliding window ([0057], lines 7-11: "For examples, stored customer information extracted from an SLA may say that the customer is entitled to the use of Z number of database servers. If the provisioning request is for Z+1, the customer is not entitled to have that provisioning request fulfilled); and

indicating application entitlement to the request for resources in response to the determining and if the request is excessive including throttling of the requested resources when the application is not entitled to the additional resources in accordance with the entitlement profile ([0057], lines 16-18: Utility computing system curtails use of resources with respect to specific time windows. "Additionally, there could be time

constraints (e.g. the customer is entitled to 20 servers between 8 am and 5 pm, but only 10 servers the rest of the time.").

Gasca fails to explicitly disclose:

"wherein the entitlement profile associated with the application describes burstiness of the application over the time period, and the burstiness includes a measure of the application's expected bursts for resources".

However, Shabauddin discloses a resource sharing system wherein predicted client behavior is modeled for purposes of forecasting future resource requirements (Shahabuddin, col. 5, lines 31-32, and col. 6, lines 10-15 and 21-24: The client resource requirement is modeled as a stationary stochastic process. Web access rates are typically distributed differently during the day. This fact is captured by dividing the day into k time slots (e.g., k=24 and each time slot is an hour long) and modeling the resource requirements of each client as a different stationary stochastic process in different time slots and for different resources. For r resources and k time slots, the resource requirement of each client is modeled as a random vector having d=r*k dimensions.). Gasca and Shabauddin are analogous art because both disclose methods for allocating shared resources in a utility computing system on a per customer basis.

It would have been obvious to one skilled in the art at the time of the invention to introduce Shabauddin's teachings of predicting client resource demands with Gasca's teachings of provisioning shard resources per customer agreement because Shabauddin would extend Gasca's provisioning system to take into account the

anticipated needs of a hosted client application. As a consequence, shared utility computing resources could be more efficiently allocated in light of predicted customer application behavior.

For claims 2 and 16, the combination of Gasca and Shabauddin discloses:
The method of claim 1 further comprising:

acquiring additional sliding windows and corresponding additional entitlement values to determine if the request for resources exceeds at least one entitlement value and sliding window combination ([0057], lines 1-4: "At block 300, the process waits for a new provisioning request". The utility computing system receives a provisioning request on behalf a customer application.); and

indicating that the application is not entitled to the requested resources when the request exceeds the entitlement value in at least one entitlement value and sliding window combination ([0057], lines 16-18: Utility computing system curtails use of resources with respect to specific time windows. "Additionally, there could be time constraints (e.g. the customer is entitled to 20 servers between 8 am and 5 pm, but only 10 servers the rest of the time.".

For claims 4 and 18, the combination of Gasca and Shabauddin discloses:
The method of claim 1 wherein a burst loading factor associated with each sliding window corresponds to the burstiness of the application and identifies a portion of an

aggregate entitlement to the resources available to fulfill the request (Shahabuddin, col. 5, lines 31-32, col. 6, lines 10-15 and col. 6, lines 21-24).

For claims 5 and 19, the combination of Gasca and Shabauddin discloses:
The method of claim 4 wherein a larger burst loading factor is associated with more bursty applications that may need resources more rapidly compared with a smaller burst loading factor is associated with applications that may not need resources as rapidly (Shahabuddin, col. 6, lines 35-45: The past access pattern of client i is used to estimate the distribution of [future client resource demands]).

For claims 6 and 20, the combination of Gasca and Shabauddin discloses:
The method of claim 1 wherein the entitlement value is derived from historical trace information collected while the application is using resources (Shahabuddin, col. 7, lines 40-44: The monitoring system 112 consequently produces a utilization pattern database 114, which is provided as input to the decision support system 116. The decision support system 116 provides suggestions to a module 122 for allocating resources optimally to clients.).

For claims 7 and 21, the combination of Gasca and Shabauddin discloses:
The method of claim 1 wherein the burst loading factor is derived from the historical trace information collected while the application is using resources (Shahabuddin, col. 7, lines 40-44).

For claims 8 and 22, the combination of Gasca and Shabauddin discloses:

The method of claim 3 wherein the resource usage is determined according to an estimated probability mass function (Shahabuddin, col. 6, lines 10-16: The client resource requirement is modeled as a stationary stochastic process.).

For claims 9 and 23, the combination of Gasca and Shabauddin discloses:

The method of claim 4 wherein the estimated probability mass function further includes a confidence interval corresponding to a sample size used for determining the estimated probability mass function (Shahabuddin, col. 6, lines 35-49: Alpha-satisfiability is introduced as a measure of QoS. A client is said to be alpha-satisfied if the client receives a promised capacity at least alpha proportion of time in each dimension.).

For claims 10 and 24, the combination of Gasca and Shabauddin discloses:

The method of claim 1 wherein the entitlement value operates as a metric for determining whether an application is entitled to the requested resources (Shahabuddin, col. 7, lines 47-49: The decision support system takes input from monitoring system produced database model and service level agreements to make decisions regarding allocation of resources).

For claims 11 and 25, the combination of Gasca and Shabauddin discloses:

The method of claim 10 wherein the entitlement value for an application is proportional to the burstiness of the application in view of resource usage derived from historical trace data (Shahabuddin, col. 7, lines 40-44).

For claims 12 and 26, the combination of Gasca and Shabauddin discloses:
The method of claim 1 wherein determining if the request for resources exceeds the entitlement value further depends on a confidence interval associated with the entitlement value and the number of sample values used to identify the entitlement value (Shahabuddin, col. 6, lines 10-15 and 35-49: A disclosure of utilization patterns having confidence interval bounds and accuracy based on number of data samples taken to construct model).

For claims 14 and 28, the combination of Gasca and Shabauddin discloses:
The method of claim 1 wherein indicating application entitlement includes clawing back resources already allocated to the application when the application has exceeded a time limit for using the allocated resources (Gasca, [0057], lines 16-18: Utility computing system curtails use of resources with respect to specific time windows. “Additionally, there could be time constraints (e.g. the customer is entitled to 20 servers between 8 am and 5 pm, but only 10 servers the rest of the time.”)

Response to Arguments

4. Applicant argues that the prior art of record does not disclose "wherein the entitlement profile associated with the application describes burstiness of the application over the time period, and the burstiness includes a measure of the application's expected bursts for resources".

Examiner disagrees. In addition to Shabauddin, col. 6, lines 10-15 and 21-24, which is cited in the rejections to the independents, Examiner also directs Applicant to Shabauddin, col. 5, line 48 – col. 6, line 10. The passage teaches a web server cluster provisioning system that provides a client web server with two kinds of resources: exclusively reserved resources and shared resources. Specifically, col. 5, line 65 - col. 6, lines 3, discloses: "Each client can be guaranteed a minimum amount of resources and also guaranteed more than the minimum with a certain probability. Since some resources are reserved exclusively for each client, a peak load of a co-hosted client does not adversely affect other clients hosted on the same shared server on the average." As such, a fair reading of the above-cited passages does in fact disclose Shabauddin's stochastic modeling process as being informed by the expected "burstiness" of resource demands. As with the instant application, Shabauddin's embodiments disclose methods for a web server farm to assign clients' web services to cluster nodes based on the clients' anticipated peak demands upon shared cluster resources.

Conclusion

5. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

a. Das et al. (20050172291): A method for utility-based dynamic resource allocation in a distributed computing system.

b. Doyle et al. (7171470): A grid service scheduling of related services using heuristics.

THIS ACTION IS MADE FINAL. Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Clayton R. Williams whose telephone number is 571-270-3801. The examiner can normally be reached on M-F (8 a.m. - 5 p.m.).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Ario Etienne can be reached on 571-272-4001. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Clayton R Williams/
Examiner, Art Unit 2457
5/19/2010

/ARIO ETIENNE/
Supervisory Patent Examiner, Art Unit 2457